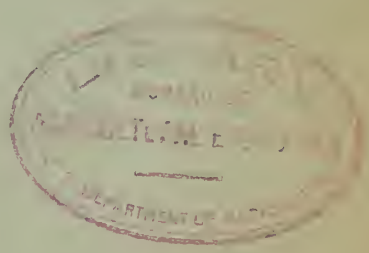
UNITED STATES DEPARTMENT OF AGRICULTURE
Agricultural Marketing Service

ANALYSIS OF VARIABILITY IN REPLICATE OBSERVATIONS

Prepared by
F. H. Harper and Norma L. Goudy



Washington, D. C.
October 1939

UNITED STATES DEPARTMENT OF AGRICULTURE
Agricultural Marketing Service

Analysis of Variability in Replicate Observations

Prepared by
F. H. Harper and Norma L. Goudy

Of special interest to workers in the Division of Cotton Marketing are procedures that have been developed and applied by the authors of this report in the analysis of cotton classing variability and in the analysis of variability in the quality of cotton ginned in the same localities from year to year. This report is concerned with the analysis and interpretation of differences in three classifications on the same press-box samples. The samples represented are a part of those procured from ginnings from the 1933 crop in Oklahoma and classed originally, along with other samples, for purposes of quality statistics assembled and disseminated in accordance with the provisions of Public No. 740 (act of Mar. 3, 1927, 44 Stat. 1372-1374).

The "x" column in table 1 contains the original observations, and the "y" and "u" columns contain other observations that were made by 2 other classers, the "y" observations having been made by 1 of these 2 classers and the "u" observations having been made by the other classer. The total number of samples, referred to in table 1 as "lots", is only 36, which is rather small for conclusive deductions, but the data will at least serve for illustrative purposes.

Variability in errors of observation and means of reducing such errors have received much attention in recent years. Analysts, particularly in England and colonial possessions of the Empire, working with agronomic

data, have evolved improved field plot technique for the purpose of reducing and minimizing the error attributable to variability in soil. More recently, the economists have manifested increasing interest in scientific analyses of factors contributing to variability in observations, as have also biologists and technologists. It is to the untiring efforts of these workers that the orderly advance in statistical procedure is largely due.

Numerous papers have been written on improvement in methods of sampling and in the analysis and interpretation of variability characterizing sample data, and many contributions have been made on the application of more logical means than formerly used for evaluating the significance of varying degrees of similarity and differences between duplicate and replicate observations. The preparation of another paper on this subject might be considered superfluous,at least to some extent, were it not for the fact that the analysis pertains to cotton, which, in certain respects, has not received as much attention and study by statistical analysts as have the cereals and a few other crops, notably potatoes and sugar beets. Furthermore, the analysis upon which this discussion is based is not concerned directly with soil heterogeneity in its relation to quantity production, but more particularly with variability in the qualitative appraisal of production.

The authors have 2 principal purposes in presenting this paper. One of these is to furnish some indication of the value and importance in certain instances of replicated observations in the elimination or reduction of error. The other purpose of the paper is to emphasize the desirability of evaluating the different parts of total variability contributed from various sources and to show how the magnitude of their differences may be interpretated.

Statistical methods that are now adaptable make it possible to separate readily the total variability into its component parts. Perhaps the most appropriate of these methods at the present time is that known as the "Analysis of Variance," 1/ suggested by Dr. R. A. Fisher, eminent statistician of the Rothamsted Experimental Station, Harpenden, England. This method of analysis is quite flexible and it can be easily applied, therefore, by making such modifications as are necessitated by the data to be analyzed and by the sources of variability detected.

The application and interpretation of the procedure and underlying fundamental principles are based in this paper on an analysis of variability in replicate series of observations on staple length of fibers in lots of cotton selected to furnish some indication of the probable predominating or characteristic length of fibers in individual bales as a whole. Emphasis is placed upon the fact that lack of agreement in the magnitude of corresponding observations in the successive series may contribute a difference between means as well as a certain range of distribution that is in addition thereto.

The term "bias" may be used for convenience to designate that part of the total variability contributed by net differences between the series

---

1/ This method was first published in preliminary form in 1923. Jour. Agri. Sci., Vol. 13, Part 3, July, 1923, pp. 311-320. A detailed explanation of the variance method of analyzing the variability in duplicate series of observations is presented in a paper entitled "The Analysis of Variance Method of Measuring Differences between the Staple-length Designations of Press-Box and Cut Samples of Cotton," by F. H. Harper and W. B. Lanham.

of observations, and the term "error" may be used in referring to the variability within the series that is estimated as being distinct and separate from both the "bias" and that part of variability attributable to actual differences among the individuals within the different series. It is obvious that the "spread" in replicate observations may contribute both "bias" and "error" because of differences between averages and because of inconsistency in variation within the different series analyzed.

Need for separately evaluating bias and error is frequently occasioned by the necessity of comparing variability contributed from different sources. It is not consistent with logical technique, of course, to isolate and express the bias and error in terms different from that representing "spread" and then attempt comparisons. Just as standard deviations calculated for distributions of entirely different kinds of data must be expressed as abstract relative measures in making comparisons of dispersion, so must the different parts of total variability be comparable if they are to be properly evaluated.

The present day analyst, fortunately, is enabled to effect these comparisons and evaluations by the application of the appropriate statistical methods. The following equations indicate the nature of calculations that may be made in deriving measures of squared variability and the fundamental principles underlying them and interpretation of results obtained from their application to the analysis of differences in three series of observations. It is obvious, of course, that these equations can be modified to provide for the analysis and interpretation of variability between and within any number of series.

1. Correction factor (c) $= \frac{[\Sigma(x + y + u)]^2}{n}$, in which "x," "y," and and "u" represent individual observations.

2. Total summation of squares, or total squared variability,

$$= \sum x^2 + \sum y^2 + \sum u^2 - c$$

3. Bias, or variability attributable to differences between series,

$$= \frac{(\sum x)^2}{n} + \frac{(\sum y)^2}{n} + \frac{(\sum u)^2}{n} - c$$

4. "Sample" variability $= \frac{\sum (x + y + u)^2}{3} - c$

5. Error = total squared variability (equation 2) - sample variability (equation 4) - bias (equation 3)

The data in table 1, of which the common mean is 30.63, are compared and the differences both between and within the series interpreted. Application of the equations is made in the interpretations, and the calculations suggested by them are explained. It will be observed that the replicate observations have not been arranged in the form of a frequency distribution. This can readily be accomplished, however, and it is a decided advantage when there is a very large number of items because such arrangement facilitates the necessary calculations. Even when there are only 36 replicate observations, such as presented in table 1, it may be advantageous in some instances to arrange them in the form of a frequency.

Table 1.- Replicate observations on the staple length of 36 lots of cotton 1/

Lot number	Staple-length designation in thirty-seconds of an inch			$\sum x+y+u$	Average of: $\frac{\sum x+y+u}{}$	$\sum x^2+y^2+u^2$
	x (observation 1)	y (observation 2)	u (observation 3)			
1	32	30	32	94	31.3	2,948
2	32	28	31	91	30.3	2,769
3	31	29	30	90	30.0	2,702
4	32	29	30	91	30.3	2,765
5	33	30	31	94	31.3	2,950
6	30	29	30	89	29.7	2,641
7	33	31	31	95	31.7	3,011
8	31	30	31	92	30.7	2,822
9	31	29	31	91	30.3	2,763
10	32	30	31	93	31.0	2,885
11	32	30	31	93	31.0	2,885
12	32	30	31	93	31.0	2,885
13	32	30	31	93	31.0	2,885
14	32	29	31	92	30.7	2,826
15	31	30	32	93	31.0	2,885
16	30	29	29	88	29.3	2,582
17	32	29	31	92	30.7	2,826
18	31	30	31	92	30.7	2,822
19	29	30	29	88	29.3	2,582
20	31	29	30	90	30.0	2,702
21	30	29	30	89	29.7	2,641
22	30	30	31	91	30.3	2,761
23	33	31	30	94	31.3	2,950
24	33	31	32	96	32.0	3,074
25	30	30	30	90	30.0	2,700
26	32	30	31	93	31.0	2,885
27	31	31	32	94	31.3	2,946
28	30	29	31	90	30.0	2,702
29	31	31	30	92	30.7	2,822
30	32	31	31	94	31.3	2,946
31	30	29	30	89	29.7	2,641
32	30	30	30	90	30.0	2,700
33	33	31	31	95	31.7	3,011
34	30	31	30	91	30.3	2,761
35	31	30	30	91	30.3	2,761
36	33	31	31	95	31.7	3,011
Total	1,128	1,076	1,104	3,308	--	101,448
Mean	31.33	29.89	30.67	30.63 2/	--	--
Sum of squares	35,388	32,184	33,876	304,122	--	--

1/ Representing observations made on cotton taken from the press box at gins in Oklahoma during the 1933-34 season.

2/ Common mean of 108 observations.

1. Correction factor $= \frac{(3,308)^2}{108}$, $= 101,322.81$

2. Total squared variability $=$ 101,448 - 101,322.81, $=$ 125.19

3. Bias = $\frac{(1128)^2}{36} + \frac{(1076)^2}{36} + \frac{(1104)^2}{36} - 101,322.81, = 37.63$

4. Sample variability = $\frac{304,122}{3} - 101,322.81, = 51.19$

5. Error = 125.19 - 51.19 - 37.63, = 36.37

Correction factor.- This value is determined by squaring the summation of all observations and then dividing by the number of observations. It represents the difference between the summation of squared observations and the summation of the squares of deviations from the common mean.

Total squared variability. The difference between the summation of squares of all observations and the correction factor is the measure of total squared variability. It is the equivalent of the total of squares of deviations of all observations from the common mean.

Bias.- This measure indicates the amount of squared variability attrib utable to differences between means of the series. It is the quantity remaining after the correction factor is subtracted from the summation of quotients obtained by dividing the squares of summation of each series by the number of observations.

"Sample" variability.- The measure of "sample" variability is obtained by squaring the summation of rows of corresponding observations in the individual series, summating the squares, dividing by 3, and then subtracting the correction factor. It represents the difference between total squared variability and the summation of those parts of the total that are attributable to bias and error, and it constitutes that part of the squared variability within the series which is in addition to that estimated as being attributable to errors of observation alone.

Error.- The measure of variability attributable to error is obtained in the calculations by subtracting from total squared variability those parts of the total that are contributed by "sample" and "bias."

In the interpretation of procedure and results it will be observed that the summation of variability within the series, 51.19 for "sample" and 36.37 for "error," is equivalent to the value obtained by summating the squares of deviations of observations in each series from its mean. It will be observed also that this summation is of the same magnitude as the difference between total squared variability and that part of the total attributable to bias, or to net differences between the series of observations as distinguished from differences within the series.

With the total squared variability separated into the designated component parts, it is then possible to proceed with the determination of whether or not one estimate of squared variability obtained from n_1 degrees of freedom 2/ differs significantly from another estimate of squared variability obtained from n_2 degrees of freedom. To do this it is only necessary to calculate the z 3/ value equal to half the difference between the natural logarithms of two derived measures of average squared variability. Then, if P represents the probability of exceeding this calculated z value by mere chance, it becomes possible to obtain the value of z corresponding to different values of P, n_1, and n_2.

2/ The term "degrees of freedom" is used in referring to the number of series and the number of observations that may be free to vary from other series and observations.

3/ The distribution of this z value is related in principle to the distribution of z values worked out by "Student" and Pearson. Fisher's z value is equal to one-half of the natural logarithm of the quotient obtained by dividing one measure of average squared variability, such as is presented in column 4 of table 2, by another. It is calculated in this report by determining the difference between one-half the natural logarithms of two derived estimates of variance, which is the equivalent of one-half the difference between two such logarithms.

The following table, based on the results of the analysis of data in table 1, is presented to clarify the procedure by which it may be determined whether or not the measure of squared variability obtained from n_1 degrees of freedom is significantly greater than that obtained from n_2 degrees of freedom.

Table 2.- Squared variability contributed from specified sources

1	2	3	4	5	6
Source of squared variability	Degrees of freedom	Squared variability (summation of squares)	Average squared variability 1/	Average squared variability times 10 2/	$\frac{1}{2}$ log.$_e$ of average squared variability times 10 3/
Bias	2	37.63	18.815	188.15	2.619
Sample	35	51.19	1.463	14.63	1.342
Error	70	36.37	.520	5.20	.824
Total	107	125.19	---	---	---

1/ Squared variability divided by degrees of freedom.
2/ Decimals moved one place to the right to avoid negative logarithms in the calculation of values in column 6.
3/ $\frac{1}{2}$ log.$_e$ equals $\frac{1}{2}$ log.$_{10}$ times 2.3026, or log.$_{10}$ times 1.1513. These values were calculated by obtaining the products of 2.3026 and one-half of the common (five-place) logarithms of the numbers in column 5.

The z value is the difference between any two of the one-half natural logarithms in column 6. Since the problem in this analysis is concerned primarily with the difference between bias and error, the desired z value is obtained by subtracting 0.824 from 2.619, which leaves 1.795. In the table 4/ showing 5 percent points of the distribution of z, with n_1 equaling 2 and n_2 equaling infinity, the z value is 0.5486, indicating that a value

4/ Fisher, R. A., Statistical Methods for Research Workers, fourth edition, 1932, table VI, pages 224 and 225. See pages 226 and 227 for 1 percent points of the distribution of z.

of z as great as or greater than 0.5486 would be expected to be obtained by chance alone in not more than 5 percent of the number of cases. With n_1 equaling 2 and n_2 equaling 60, the z value occurring at the 5 percent point is 0.5738. In the table of 1 percent points of the distribution of z, with n_1 equaling 2 and n_2 equaling infinity, the z value is 0.7636. The logical conclusion seems to be, therefore, that the calculated z value of 1.795 indicates the difference between the two variances, bias and error, to be quite significant.

Of perhaps equal importance is the magnitude of the two measures of squared variability for bias and error, presented in column 3 of table 2. It will be observed that the degrees of freedom for bias and sample are 2 and 35, respectively, and it will be observed further, by reference to column 4, that the average squared variability for error is 0.520. When the appropriate parts of this measure for error are subtracted from the squared measures in column 3 for bias and sample, and these subtracted parts are added to the squared measure for error in column 3, it is then possible to express these squared measures in terms of percentages of the total. This procedure may be convenient when it is desirable to obtain information on the proportionate contributions made from the various sources to the total summation of squares.

On Measuring Proportionate Contributions to Total Variability

One of the many interesting statistical procedures evolved and described by Dr. W. B. Kemp is that on "Some Methods for Statistical Analysis" published in the June, 1934, issue of the Journal of the American Statistical Association, Vol. XXIX, No. 186, pp. 147-158. In that paper,

the discussion is devoted in large measure to a treatment of ways of developing and interpreting known relationships, and there is presented an easily understood analytical approach that is based on methods and principles underlying stratification processes.

It is gratifying to see this paper in print. The general procedure described has been taught by Dr. Kemp in his graduate classes, and many phases of it have been put to practical use since first described by him. The authors of this note have found particularly valuable that part of the procedure which provides for the separation of total variability into its component parts free from estimated error accompanying them.

By this method it has been found possible, after determining total squared variability in paired and replicate series, to isolate the component parts, free from estimated error, contributed from different detected sources, and then to express these component parts as percentages of the total. It is thus possible to appraise readily the relative magnitudes of different component parts of variability. And, as Dr. Kemp has explained, the procedure advanced by him, whereby variability can be analyzed into its component parts free from estimated error, permits a ready transfer to the "Analysis of Variance," which gives results in which error is contained in the individual component parts.

Note on the Revised Cotton Grade Standards

that Became Effective August 20, 1936

In the revision of the grade standards for upland cotton which was promulgated in 1935, effective August 20, 1936, the grades for Blue Stained and Light Yellow Stained cotton were eliminated and the grades for Yellow Stained were made descriptive. The grades for White cotton were shifted to include whiter color and the more creamy bales in the higher grades for White cotton were eliminated, since creamy colored bales could not be found in quantities sufficient to make copies of the old standards. The highest grade for White cotton, No. 1 or Middling Fair, was made descriptive, as were the grades for Extra White. The latter were increased in number from 5 to 7. The new grades for Tinged cotton do not contain the deeper colored cottons of the old series and the changes in the White and in the Tinged standards resulted in excluding from the descriptive standards for Spotted cotton much of the Light Tinged cotton heretofore classified as Spotted.